

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 19, 2002, 07:53:17 ; Search time 1603 Seconds  
(without alignments)  
3856.273 Million cell updates/sec

Title: US-09-807-459-2  
Perfect score: 2359  
Sequence: 1 MAPSDVGDVTKTLAASES:.....DPSKALIRKVTSEADNLEK 458

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 13736207 segs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ -p2n model -DEV-xlh  
-Q/cgn\_1/OSPT0.spool/US09807459/runat\_18102002.141112\_28595/app-query.fasta.1.647  
-DB-EST -QFMT-fastlap -SUFFIX-rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS-bits -FASTAP -END=-1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45  
-DOCLIN=200 -THR SCORE-pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE-LOCAL  
-OUTFMT-pio -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09807459\_ECGN\_1.1\_753\_etunat\_18102002.141112\_28595 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAR -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEDOUT=120  
-WARR\_TIMEDOUT=30 -THREADD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:.\*  
2: em\_esthum:.\*  
3: em\_estln:.\*  
4: em\_estmu:.\*  
5: em\_estov:.\*  
6: em\_estpl:.\*  
7: em\_estro:.\*  
8: em\_hlc:.\*  
9: gb\_estl:.\*  
10: gb\_est2:.\*  
11: gb\_hlc:.\*  
12: gb\_gss:.\*  
13: em\_gss\_hum:.\*  
14: em\_gss\_inv:.\*  
15: em\_gss\_pln:.\*  
16: em\_gss\_vit:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	131	5.6	529	12	A2167872
2	123	5.2	564	10	BM275292

## SUMMARIES

C	3	120	5.1	952	12	A2209422	A2209422 SP_0106_A
C	4	118	5.0	856	12	A2199582	A2199582 SP_1033_B
C	5	117.5	5.0	1035	10	BM416596	BM416596 Op21695 M
C	6	117	5.0	504	10	BM274689	BM274689 PFESTroa7
C	7	117	5.0	537	10	BM274676	BM274676 PFESTroa7
C	8	115	4.9	366	12	BM2140870	BM2140870 SP_0001_B
C	9	113	4.8	599	12	A2199581	A2199581 SP_1039_B
C	10	113	4.8	665	12	A2400717	A2400717 IM0167E11
C	11	112.5	4.8	897	12	CNS02TD6	AL213027 Tetradon
C	12	112	4.7	624	9	AA571652	AA571652 vm09c03_r
C	13	111	4.7	794	12	BH098592	BH098592 RPCT-24-2
C	14	110.5	4.7	698	12	A2766807	A2766807 IM0564C11
C	15	109	4.6	324	12	A2152840	A2152840 SP_0034_B
C	16	109	4.6	642	12	A2988471	A2988471 2M027021
C	17	109	4.6	700	9	BB230024	BB230024 BB230024
C	18	109	4.6	943	11	BC019933	BC019933 Mus muscu
C	19	108.5	4.6	721	12	A2818810	A2818810 2M0082P07
C	20	108	4.6	996	12	CNS06H10	AL439038 T3 end of
C	21	108	4.6	1154	12	CNS07C18	AL438594 T3 end of
C	22	107.5	4.6	642	12	A2431579	A2431579 IM0216006
C	23	106	4.5	640	12	A2733090	A2733090 RPCT-24-1
C	24	105	4.5	478	12	A2462925	A2462925 IM0271A02
C	25	104.5	4.4	532	10	BM273998	BM273998 PFESTroa6
C	26	104	4.4	596	12	A2944493	A2944493 2M0205C06
C	27	104	4.4	644	10	BM168451	BM168451 EST570974
C	28	103.5	4.4	715	12	BH126086	BH126086 RPCT-24-2
C	29	103	4.4	717	10	BG523384	BG523384 31-25 Ste
C	30	103	4.4	791	12	BH047254	BH047254 RPCT-24-2
C	31	102.5	4.3	439	12	A2946473	A2946473 2M0208B13
C	32	102.5	4.3	503	12	A2344512	A2344512 IM0078M05
C	33	102.5	4.3	547	12	A2696548	A2696548 RPCT-23-2
C	34	102	4.3	869	12	A2679969	A2679969 ENTEZ06TR
C	35	101.5	4.3	683	12	BH070091	BH070091 RPCT-24-2
C	36	101.5	4.3	830	12	A2667704	A2667704 ENTKT09TF
C	37	101	4.3	605	12	A2891746	A2891746 RPCT-24-2
C	38	101	4.3	614	12	A2415898	A2415898 IM0190A17
C	39	101	4.3	622	12	A2625698	A2625698 IM0465N20
C	40	101	4.3	700	12	BH313240	BH313240 CH230-101
C	41	100.5	4.3	613	12	A2342112	A2342112 IM0075F05
C	42	100.5	4.3	651	12	BH352781	BH352781 CH230-175
C	43	100.5	4.3	658	12	A2421595	A2421595 IM0199K23
C	44	100.5	4.3	710	12	BH278149	BH278149 CH230-48P
C	45	100.5	4.3	739	12	BH036938	BH036938 RPCT-24-3

## ALIGNMENTS

RESULT 1	A2167872	529 bp	DNA	linear	GSS 29-AUG-2000
LOCUS	SP_0103_A2_E10_77A	Strongylocentrotus purpuratus			
DEFINITION	, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=103 Col=20 Row=I, DNA sequence.				
ACCESSION	A2167872				
VERSION	A2167872.1	GI:8338240			
KEYWORDS	GSS.				
SOURCE	Strongylocentrotus purpuratus.				
ORGANISM	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae; Strongylocentrotus.				
REFERENCE	1 (bases 1 to 529)				
AUTHORS	Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Bondi, T.R., Swartzell, S., Wallace, J.C., Pousta, A.J., Livingston, B.T., Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and Hood, L.				
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)				
MEDLINE	20402566				
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology				



DB 184 CCAACAAATGACCTTAAATGACCAATGACCAATGACCAATGACCAATGACCAATGAC 243  
 QY 366 GlllleArgAspProSerlysalaleuilearglyvalserthrlelyalagluaspleu 365  
 DB 244 CCAATTAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCA 282  
 QY 386 PhegluasnllylelglynglythrValAspPheIleasnngluileargaspPro 405  
 DB 283 ATTAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGAC 342  
 QY 406 Serlysalaleuilearglyvaltyrrhglualaspaspleuphegluasnllyle 425  
 DB 343 ATTAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGAC 393  
 QY 426 GlylgnglythrValAspPheIleasnngluileargaspPro 440  
 DB 394 AATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCA 438

RESULT 3  
 AZ209422/c 952 bp DNA linear GSS 31-AUG-2000  
 LOCUS SP\_0106\_A2\_A12-SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
 DEFINITION genomic clone Plate=106 Col=24 Row=A, DNA sequence.  
 AZ209422  
 ACCESSION AZ209422.1 GI:8423111  
 KEYWORDS GSS.  
 SOURCE Strongylocentrotus purpuratus.  
 ORGANISM Strongylocentrotus purpuratus.  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Euechinoidea; Echinoidea; Echinoidea;  
 Strongylocentrotidae; Strongylocentrotus.  
 1 (bases 1 to 952)  
 Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,  
 Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray,  
 G.A., Eitensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and  
 Hood, L.  
 A sea urchin genome project: Sequence scan, virtual map, and  
 additional resources  
 Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
 20402566  
 CONTACT Cameron, R.A., Davidson, E.H., Hood, L.  
 DIVISION Division of Biology 156-29  
 California Institute of Technology  
 Pasadena California 91125, USA  
 Tel: (626) 395-8421  
 Fax: (626) 793-3047  
 Email: acameron@caltech.edu  
 Plate: 106 Row: A Column: 24  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 952.  
 Location/Qualifiers  
 1..952  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone="Plate=106 Col=24 Row=A"  
 /clone\_lib="Strongylocentrotus purpuratus, purple sea  
 urchin, sperm genomic BAC library"  
 /note="Organ: sperm; Vector: BAC#3, 6; BAC Clones in E-Coli  
 DH10B"

BASE COUNT 145 a 265 c 270 g 272 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.00381 Length: 952  
 Score: 120.00 Matches: 38  
 Percent Similarity: 42.57% Conservative: 25  
 Best Local Similarity: 25.68% Mismatches: 69  
 Query Match: 5.09% Indels: 16  
 Gaps: 12 4

US-09-807-459-2 (1-458) x AZ209422 (1-952)

QY 294 Asnlysgluilearglyvalserlysalaleuilearglyvalserthrlysal 313  
 DB 523 AACAGGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGAC 464  
 QY 314 AspleuphegluasnllylelglynglythrValAspPheIleasnngluilearg 333  
 DB 463 GAATCATGAGAGAACCGGATTTCTTGGGGGAC-----AACCAAGCATTCGAG 416  
 QY 334 AspProSerlysalaleuilearglyvalserasnaspalelyasphegluasn 353  
 DB 415 ACTCAAGTCAGCGCTGCGCGCTGCTGCGATGAGACCGAGATGATGAGAAC 356  
 QY 354 LysIlelglynglythrValAspPheIleasnngluileargaspProSerlysa 373  
 DB 355 CGGTATTCCTTGGGGAT-----AACCAAGCATTCGAGACTCCAGCTCCAGCC 308  
 QY 374 Leuilearglyvalserthrlysalaleuilearglyvalserlysalaleuilearg 393  
 DB 307 TCGGCGGTCTCTGCGAGGAGACCGAGATTCGATGAGAACCGGATTTCTTGGGG 248  
 QY 394 ThrValAspPheIleasnngluileargaspProSerlysalaleuilearglyval 413  
 DB 247 GCC-----AACCAAGCATTCGAGACTCCAGCTCCAGCTCGCGCTGCTG 200  
 QY 414 TyrThrGluAlaspaspleuphegluasnllylelglynglythrValAspPhe 433  
 DB 199 GCATGGACACCGCATTCGATGAGAACCGGATTCCTTGGGGAT----- 152  
 QY 434 AsnlysgluileargaspProSer 441  
 DB 151 AACCAAGCATTCGAGACTCCAGCT 128

RESULT 4  
 AZ199582/c 856 bp DNA linear GSS 31-AUG-2000  
 LOCUS SP\_1039\_B2\_E12-77A Strongylocentrotus purpuratus, purple sea urchin  
 DEFINITION , sperm genomic BAC library Strongylocentrotus purpuratus genomic  
 clone Plate=1039 Col=24 Row=J, DNA sequence.  
 AZ199582  
 ACCESSION AZ199582.1 GI:8394390  
 KEYWORDS GSS.  
 SOURCE Strongylocentrotus purpuratus.  
 ORGANISM Strongylocentrotus purpuratus.  
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 Echinoidea; Euechinoidea; Echinoidea; Echinoidea;  
 Strongylocentrotidae; Strongylocentrotus.  
 1 (bases 1 to 856)  
 Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,  
 Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray,  
 G.A., Eitensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and  
 Hood, L.  
 A sea urchin genome project: Sequence scan, virtual map, and  
 additional resources  
 Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
 20402566  
 CONTACT Cameron, R.A., Davidson, E.H., Hood, L.  
 DIVISION Division of Biology 156-29  
 California Institute of Technology  
 Pasadena California 91125, USA  
 Tel: (626) 395-8421  
 Fax: (626) 793-3047  
 Email: acameron@caltech.edu  
 Plate: 1039 Row: J Column: 24  
 Seq primer: 17  
 Class: BAC ends  
 High quality sequence stop: 856.  
 Location/Qualifiers  
 1..856  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone="Plate=1039 Col=24 Row=J"

FEATURES  
 source  
 1..856  
 /organism="Strongylocentrotus purpuratus"  
 /db\_xref="taxon:7668"  
 /clone="Plate=1039 Col=24 Row=J"



```

DEFINITION  PFESToa73e05.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
Plasmodium falciparum 3D7 cDNA 5' similar to TR:097255 097255
MAL3P3.1 PROTEIN.; mRNA sequence.
ACCESSION   BM274689
VERSION     BM274689.1
KEYWORDS    GI:17968000
SOURCE      EST.
ORGANISM    Plasmodium falciparum 3D7.
SOURCE      Plasmodium falciparum 3D7.
REFERENCE   1 (bases 1 to 504)
AUTHORS     Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Marras,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagarisshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Ritchey,J., Wadkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
JOURNAL     Unpublished (2001)
COMMENT     Washu Plasmodium EST Project
Contact: L. David Sibley
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by R. Haywood. DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -40up from Gibco
High quality sequence stop: 416.
FEATURES
    source
        1..504
        /organism="Plasmodium falciparum 3D7"
        /db_xref="taxon:36329"
        /clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
        library"
        /dev_stage="gametocyte (stage III-V)"
        /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
        /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
        XhoI. The library was constructed by R Haywood. cDNAs were
        synthesized from gametocyte poly(A)+ RNA by oligo d(T)
        priming, size-selected and directionally cloned into the
        EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR
        lambda vector (Stratagene). The primary library was mass
        excised as phagemid using the ExAssist helper phage
        (Stratagene). Clones were mass excised using the ExAssist
        helper phage (Stratagene), the phagemids were precipitated
        with PEG 8000 and extracted with phenol/chloroform.
        Phagemid DNA was electroporated into DH10B cells. Clone
        Availability: David Sibley, Washington University."
BASE COUNT  253 a      130 c      42 g      78 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      0.00316      length:      504
Score:          117.00      matches:      41
Percent Similarity: 38.71%      Conservative: 19
Best Local Similarity: 26.45%      Mismatches: 81
Query Match:    4.96%      Indels:      14
DB:            10      Gaps:      4
US-09-807-459-2 (1-458) x BM274689 (1-504)
OY 287 ILEProthLyLysPheAsnLysGluIleArgLupProSerLysAlaLeuLysGlu 306
      ::::::::::: ||| :::::::::::
DB 13 GTCACAAATGACCAATTAATGACCAATTAATGACCA-----ATAATGTC 63
OY 307 LysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValaap 326
      ::::::::::: ||| :::::::::::
DB 64 CCAATTAATGACCAATTAATGACCAATTAATGACCAATTAATGACCAATTAATGAC 123
OY 327 PhePheAsnLysGluIleArgAspPro---SerLysAlaLeuLysGluLysValSerAsn 345
DEFINITION  PFESToa73e03.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
Plasmodium falciparum 3D7 cDNA 5' similar to TR:097255 097255
MAL3P3.1 PROTEIN.; mRNA sequence.
ACCESSION   BM274676
VERSION     BM274676.1
KEYWORDS    GI:17967987
SOURCE      EST.
ORGANISM    Plasmodium falciparum 3D7.
SOURCE      Plasmodium falciparum 3D7.
REFERENCE   1 (bases 1 to 537)
AUTHORS     Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Marras,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagarisshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Ritchey,J., Wadkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
JOURNAL     Unpublished (2001)
COMMENT     Contact: L. David Sibley
Washu Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by R. Haywood. DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -40up from Gibco
High quality sequence stop: 423.
FEATURES
    source
        1..537
        /organism="Plasmodium falciparum 3D7"
        /db_xref="taxon:36329"
        /clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
        library"
        /dev_stage="gametocyte (stage III-V)"
        /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
        /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
        XhoI. The library was constructed by R Haywood. cDNAs were
        synthesized from gametocyte poly(A)+ RNA by oligo d(T)
        priming, size-selected and directionally cloned into the
        EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR
        lambda vector (Stratagene). The primary library was mass
        excised as phagemid using the ExAssist helper phage
        (Stratagene). Clones were mass excised using the ExAssist

```



REFERENCE	Strongylocentrotidae; Strongylocentrotus.
AUTHORS	1 (bases 1 to 599) Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartell,S., Wallace,J.C., Pouska,A.J., Livingston,B.T., Wray G.A., Etlensohn,C.A., Lehnach,H., Britten,R.J., Davidson,E.H. and Hood,L. A sea urchin genome project: Sequence scan, virtual map, and additional resources Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
JOURNAL MEDLINE COMMENT	20402366 Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 1039 row: J column: 24 Seq primer: SP6 Class: BAC ends High quality sequence stop: 599.
FEATURES	location/Qualifiers 1..599 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone_plate="1039 Col=24 Row=J" /clone_id="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BAC63.6; BAC clones in E-Coli DH10B"
BASE COUNT	154 a 168 c 167 g 110 t
ORIGIN	
Alignment Scores:	
Pred. NO.:	0.0121 Length: 599
Score:	113.00 Matches: 30
Percent Similarity:	45.22% Conservative: 22
Best Local Similarity:	26.09% Mismatch: 51
Query Match:	4.79% Indels: 12
DB:	Gaps: 3
US-09-807-459-2 (1-458) x AZ199581 (1-599)	
OY	294 AsnLysGluIleArgGluProSerLysAlaLeuLysGluLysValSerThrAspPheLys 313                 :   :    :    :    : Db 11 AACCAAGCATCGACATCCAAAGTCACGCTGGCGGTCTCTGCATGGACAGCCGA 70 OY 314 AspLeuPheGluAsnLysLysIleGlyInclgLythrValAspPheAsnLysGluIleArg 333                 :   :    :    :    : Db 71 GATTTCGATGGAGAACCGGTATTCTTTGGGGGAT-----AACCAAGCATCCAAAG 118 OY 334 AspProSerLysAlaLeuLysGluLysValSerAsnAlaPalatAspLeuPheGluLys 353                 :   :    :    :    : Db 119 ACTCCAATCAACCTCGGCCGCTGCTCTGGACGGGACAGCCGAGAATCGATGGAGAAC 178 OY 354 LysIleGlyInclgLythrValAspPheIleAsnAngluIleArgAspProSerLysAla 373 :::         :   :    :    :    : Db 179 CGGTATTCTTGGGGGAC-----AACCAAGCATCAAAAACCCCAAGTCCAGCC 226 OY 374 LeuIleArgLysValSerThrGlyAlaGluLysPheGluAsnLysIleGlyInclgLy 393 ::::: ::  :   :               Db 227 TCGGCCGTGCTCTGGCATGGACAGCCGACAGCAAGCAAGAGAACCAATATTCTTGGGG 286 OY 394 ThrValAspPheIleAsnAngluIleArgAspProSerLysAla 408                   :   :    :    : Db 287 GAT-----AACCAAGCATCGAAGACCCCAAGTCCAGC 319
RESULT 10	
LOCUS	AZ400717 665 bp DNA linear GSS 03-OCT-2000
DEFINITION	IM0167E1F Mouse 10kb plasmid UUCGCM library Mus musculus genomic
ACCSSION	AZ400717 clone UUCGCM10167E1 F, DNA sequence.

VERSION	AZ400717.1	GI:10515791
KEYWORDS	GSS.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 665) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D., Weiss,R.	
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts	
JOURNAL	unpublished (2000)	
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0167 row: E column: 11 Seq primer: CGTGTAAACGACGCCACGT Class: plasmid ends High quality sequence stop: 665. Location/Qualifiers 1..665 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="U06C1M0167E11" /clone_lib="Mouse 10kb plasmid U06C1M library" /sex="Male" /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-" /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g14732141g149b1f29072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."	
BASE COUNT	301 a 132 c 122 g 110 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	0.0142	Length: 665
Score:	113.00	Matches: 58
Percent Similarity:	41.94%	Conservative: 46
Best local Similarity:	23.39%	Mismatches: 76
Query Match:	4.79%	Indels: 68
DB:	12	Gaps: 13
US-09-807-459-2 (1-458) x AZ400717 (1-665)		
OY 222	SerArgLauGlnHisIleThrSerSerIyrLysAspIyrMetAspThrGlnIleProAla	241
Db 21	GCAAACGTTAAGAACTTCTTACACGAACCAAGGCCACTCACCATCAGACAAACCCAGC	80
OY 242	LeuProlLysPheAlaLysArgPheSerLeuIleValIaIaGlnArgLeuAlaIarThVal	261

```

Db      81  ATTCCTCAACTAACCACT-----
Qy      262  ALGlyTyrValAspThrProTyrLysTyrIleMetLysLeuLysAsnPheMet 281
          |||||
Db      99  CCGGATACCCCAACACACCTGCGATTATTAACATATCTC 152
          |||||
Qy      282  ValAsnArgValPheIleProThrLysLysPheAsn-----LysGluIleArgGlu 299
          |||||
Db      153  ATGATGCTGTAGAGCACTTTAAGAAAGCATTTATTAACCTTAAGAAATACAGAG 212
          |||||
Qy      300  ProSer-----LysAlaLeuLysGluLysValSerThrAspThrLysAspLeu 315
          |||||
Db      213  AACACTGCCAAAAGAGTAGAGTGTAAAGAAATATAGAAAAACACACACAGGTG 272
          |||||
Qy      316  PheGluAsnLysIleGlyGlnGlyThrValAspPheAsnLysGluIleArgAspPro 335
          |||||
Db      273  ATGGAA-----TTTAACAAAACCTTCACAGACCTA 302
          |||||
Qy      336  SerLysAlaLeuLys-----GluLysValSerAsnAspAlaLys-----AspLeu 350
          |||||
Db      303  AAAAGGAGAGTAGAAATATAGAAAAACGAAAGTGAACACTGTGAGATAGAAAC 362
          |||||
Qy      351  PheGluAsnLysIleGlyGlnGlyThrValAsp---PheIleAsnAsnGluIleArgAsp 369
          |||||
Db      363  CTAGAAGATATA-----TCAGAAACCATAGATGCAATCATCAGTAAACAGATACAGAG 416
          |||||
Qy      370  ProSerLysAlaLeuIleArgLysValSerThrGlyLysGluAspLeuPheGluAsn--- 388
          |||||
Db      417  ATGGACAGAGCAATC-----TCAGTCTCAGAAAGTTCCAAGACAGACATG 461
          |||||
Qy      389  -----LysIleGlyGlnGlyThrValAspPhe 397
          |||||
Db      462  GGCACACCAATCAAAAGAAATGCAAAATGAAAAAGATCCTAATCTCAAAATATCCAGAA 521
          |||||
Qy      398  IleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGluAla 417
          |||||
Db      522  GCCCAGGACACATGAGAGACCAAC-----CTACGATATATGCGATGAGAG 572
          |||||
Qy      418  AspAspLeuPheGlu-----AsnLysIleGlyGlnGlyThr 429
          |||||
Db      573  AATGAAGATTTTCAACTTAAGGGCCAGCAAAATCTTCAACAAATATATAGACGGGAAAC 632
          |||||
Qy      430  ValAspPheIleAsnLysGluIle 437
          |||||
Db      633  TTCCCATACCTTAAGAAAGAGATG 656
          |||||

RESULT 11
LOCUS      CNS02TD6               897 bp    DNA    linear    GSS 15-MAY-2000
DEFINITION Tetradodon nigroviridis genome survey sequence PUC-ori end of clone
              16314 of library G from Tetradodon nigroviridis, genomic survey
              sequence.
ACCESSION   AL213027
VERSION     AL213027.1 GI:7871846
KEYWORDS    GSS: genome survey sequence.
SOURCE      Tetradodon nigroviridis.
ORGANISM    Tetradodon nigroviridis.
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
              Tetraodontidae; Tetradodon.
REFERENCE   1 (bases 1 to 897)
AUTHORS     Roest-Crolius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C.,
              Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
              Weissenbach,J.
TITLE       Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetradodon nigroviridis
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 897)
AUTHORS     Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
              Bernot,A., Fitzames,C., Winkler,P., Brotlier,P., Quetier,F.,
              Saurin,W. and Weissenbach,J.

```

```

TITLE      Human gene number estimate provided by genome wide analysis using
            Tetradodon nigroviridis DNA sequence
REFERENCE   3 (bases 1 to 897)
AUTHORS     Unpublished
            Genoscope.
JOURNAL     Direct Submission
COMMENT      Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
            This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetradodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/tetradodon.

FEATURES             source
     source           1..897
                     /organism="Tetradodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="16314"
                     /clone_11b="G"
                     /note="Genoscope sequence ID : COAG163DE07SP1-end :
                     PUC-ori"

BASE COUNT      301 a      179 c      244 g      161 t      12 others
ORIGIN
Alignment Scores:
Pred. No.:      0.0258      Length:      897
Score:          112.50      Matches:      77
Percent Similarity: 40.81%      Conservative: 54
Best Local Similarity: 23.99%      Mismatches: 100
Query Match:    4.77%      Indels:      90
DB:             12      Gaps:      18

US-09-807-459-2 (1-458) x CNS02TD6 (1-897)
Qy      192  ThrGlyLeuPheGlyTyrPglyIleLysArgAlaLeuLysGln-----IleIleArg 208
          |||
Db      29  ACTGCACATTTGGATGGGAGTAGTATTCCTCGATGATGAGGGCTATTGAGTCTATG 88
          |||||
Qy      209  SerAsnLeuPro-LeuAspIleGlyThrGluHisSerValSer----- 222
          |||
Db      89  TTCACCTCTCCACATTCGACTGTGATCTGACGAGAGTAGTCCCTCGAGAAAGAACCC 148
          |||||
Qy      223  -----ArgLeuGlnHisIleThrSerSerTy 231
          |||||
Db      149  TGAATCTCGAAAAATTTGATGTCACACTTGTTCAGAAACTCGAGAAATTTGAAGCTTGCT 208
          |||||
Qy      231  rLysAspTyr---MetAspThrGlnIleProAla-----LeuProLy 244
          |||||
Db      209  GACGAAATTATCAGCAGGACTTGACCTTCGACATCGAGAGATTACAGCATCTGCTGAA 268
          |||||
Qy      244  sPhe-----AlaLysArgPheSerLeuMetValVal----- 254
          |||||
Db      269  ACTAGAGACGACGAAAGGAGAGTAGTCAGACGCGAGAGAGAGACCAACGATCAGATGCG 328
          |||||
Qy      255  -----GlnArgLeuLeuAlaThrValAlaGlyTyrValAspThrProTyrTyrLy 271
          |||||
Db      329  TGCAGAAATATCAGCCTCTCTGACAGCATCGCC-----GACGTCCAAAACACAGA 379
          |||||
Qy      271  sLysTyrPThrLysLysLysAsnPheMetValAsnArgValPheIleProThrLysLy 291
          |||||
Db      380  GAAACAGCTGGAAGAAACCTCAACGCTGTGACAGAGAACATGATG----- 422
          |||||
Qy      291  sPhePheAsnLysGluIleArgLysGluProSerLysAlaLeuLysGluLysVal----- 308
          |||||
Db      423  -----GGAAACACACACGAGAGAGACAGATCCCTCCAGAAAGAAAGAACACAGA 472
          |||||
Qy      309  SerThrAspThrLysAspLeuPheGluAsnLys-----IleGl 321
          |||||
Db      473  CGAGAACGCGTCAGACACTTATCCCGTCAGCAGACAGATTKACAMTTTATATCA 532
          |||||
Qy      321  yGlnGlyThrValAspPhePheAsnLysGluIleArgAspPro---SerLysAlaLeuLy 340
          |||||
Db      533  AAAGACCGACGTCGAGGAGAGAGAAAGAAATACAGAGAGAGAGCTCACTCACTCTAAA 592
          |||||
Qy      340  sGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThr-- 359
          |||||

```



```

|||||:-----||| :|||:-----||| :|||
593 AGAGAGATG-----GACATGAGGACATGAGGAAGTCACAGCCAGCAGCAG 646
OY 360 -----Valaspheileasnsnlglnleargasproserlysalauleuilearglysta 378
Db 647 GATGAAGATTTC-----AGCAATTACTCAAC-----CTGATAAGGAACT 688
OY 378 lserthrglyalagluaspheugluasnlyslleglynglnlythValaspheile 398
Db 689 CGAGCAGCAGGTGAGCAGCTC-----CAGGACTCGGCAAGATGAM 730
OY 388 easnnsnlglnleargasproserlysalauleuilearglystaValytrthcluilaas 418
Db 731 TAAACACAGGAAGAAAGACAGATTCTCCCTGTGAAAAAATTTGAAATTTCTTC-- 788
OY 418 psaspheugluasnlyslleglynglnlythValaspheileasnlysluilear 438
Db 789 ----CTTTCAGATCAGAACATTAAGGGAATGTGATTTCTCATCAGAGAGAAAG 844
OY 438 gasproserlysalauleuilearglystaValserthrgluaspheileasnlysluile 458
Db 845 GGAG-----CTGAAGAAAAAGCCGACACCTGGAGAGAAAG 880
OY 458 s 458
Db 881 A 881

```

## RESULT 12

AA571652

## LOCUS

624 bp mRNA linear EST 27-AUG-1997  
 vm9c03.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA  
 clone IMAGE:989668 5' similar to gb:U15647\_cds1 Mus musculus (mouse  
 ):: mRNA sequence.

## ACCESSION

AA571652

## VERSION

AA571652.1 GI:2346581

## KEYWORDS

EST.

## SOURCE

house mouse.  
Mus musculus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

Matta, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

## TITLE

The WashU-HMI Mouse EST Project

## JOURNAL

Unpublished (1996)

## COMMENT

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG1:361948  
High quality sequence stop: 469.

## FEATURES

source

1..624

Location/Qualifiers

/organism="Mus musculus"

/strain="B6D F1/J"

/db\_xref="taxon:10090"

/clone="IMAGE:989668"

/clone\_lib="Knowles Solter mouse blastocyst B1"

/issue\_type="blastocyst"

/dev\_stage="embryo (pre-implantation)"

/lab\_host="DH10B"

/note="Organ: embryo; Vector: pSPORT; Site: 1; NotI;  
Site 2: SalI; Cloned unidirectionally from mRNA prepared  
from 800 blastocysts. Primer: SalI(4T):  
5'-CGTTCGACCGTCGACCGCTTTTCTTTTCTTTT-3', CDNA5 were

cloned into the NotI/SalI sites of a pSPORT vector (Life  
 Technologies). Two different size selections: B1 (larger  
 inserts) and B3.

## BASE COUNT

293 a 117 c 114 g 100 t

## ORIGIN

Alignment Scores:

Score: 0.0168  
 Percent Similarity: 112.00  
 Best Local Similarity: 43.908  
 Query Match: 30.49%

Mismatches: 50  
 Indels: 60  
 Gaps: 32

US-09-807-459-2 (1-458) x AA571652 (1-624)

OY 295 Lysgluileargluarproser-----LysalauleuylsgluysValserth 310  
 Db 75 AAGAAATACAGAGAACAGAGTAAGACAGCTGAAGCCCTTAAGAGAAACACAA 134  
 OY 311 AspThlyaspheugluasnlyslleglynglnlythValaspheileasnly 330  
 Db 135 TCCCTTAAGAGATTACAGGAACACATCAACAG-----GTGAAGAAATGAAC 188  
 OY 331 Gluileargasproserlysalauleuylsglu-----LysValserasnaspala 349  
 Db 189 ACCATCCAGATCTTAAATGCACTAGAACATTAAGAAATACCAAGGAGACAC 248  
 OY 350 Leuphegluasnlyslleglyngln-----GlyThrValaspheileasnnglu 366  
 Db 249 CTGCAATTGAAGAAACCTAGAAAGAGATCGAGTCAATGATCAGACATCACCAC 308  
 OY 367 lIeargasproserlysalauleuilearglystaValserthrglyalagluasphe 386  
 Db 309 ATCAAGAC-----ATAGAAGAGAGATCTCAAGGCTCGGAAATATACCAT 356  
 OY 387 GluasnlyslleglynglnlythValaspheileasnngluileargasproser 406  
 Db 357 GAAAC-----ATTGACCAACAGTCAAGAAATGCAAAACCAA 398  
 OY 407 LysalauleuilearglystaValytrthgluilaasp----- 418  
 Db 399 AACCTCTTAACCAACATCTGGAATTCAGACACATGAGAACCAACCTAAG 458  
 OY 419 -----AspIeuphegluasnlyslleglyngln-----GlyThrValaspheileasn 434  
 Db 459 ATATATATATATAGAGAGATTCACACTTAAGAGCCAGTAAATATCTTCAAC 518  
 OY 435 Lysgluilearg 438  
 Db 519 AAAAATTATAGA 530

## RESULT 13

BH098592

## LOCUS

BH098592

## DEFINITION

794 bp DNA linear GSS 19-JUL-2001  
 RPI-24-287H12-TV RPI-24 Mus musculus genomic clone RPI-24-287H12  
 , DNA sequence.

## ACCESSION

BH098592

## VERSION

BH098592.1 GI:14920536

## KEYWORDS

GSS.

## SOURCE

house mouse.  
Mus musculus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 794)

## AUTHORS

Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M.,  
Tsegaye, G., Geier, K., Kroll, M., Shvartsbeyn, A., Gebregorgis, E.,  
Russell, D., de Jong, P. and Fraser, C.M.

## TITLE

Mouse BAC End Sequences from Library RPI-24

## JOURNAL

Unpublished (1999)

## COMMENT

Other GSSs: RPI-24-287H12-TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics



of pMD2.1 (14732114.19b/Plasmid25072.1), a copy-number inducible derivative of pAS12072.1). The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT	335 a	120 c	125 g	118 t
ORIGIN				

**Alignment Scores:**

Pred. No.:	0.0298	Length:	698
Score:	110.50	Matches:	50
Percent Similarity:	45.70%	Conservative:	35
Best Local Similarity:	26.88%	Mismatches:	56
Query Match:	4.68%	Indels:	5
DB:	12	Gaps:	11
			15

US-09-807-459-2 (1-458) x A2766807 (1-698)

Oy	277	LeuYsaasNPheMcValAsnArgValPheIleProthrlYsLysPhePheAsn-----	294
Db	6	TTAAATCATATCTCATACATGCTAGAAAGACATTAAGAAGAGACTTTAATTAACACTT	65
Oy	295	LysGluIleIeaIgluProSer-----LysAlaLeuYsGluYsValSerThr	310
Db	66	AAAGAAATATACGGAGAACCTCGCTPAAAGAGTTTACAAAGTCTCTTAAAGAAAAACAGGAAAC	125
Oy	311	AspThrIysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLys	330
Db	126	ACATCTCAAAAGAGCTGATGGAA-----ATGAACAAA	155

Oy	331	GluIleatgaaspProSerLysAlaLeuLys-----GluLysValSerasnAspaLaLys	348
		::::  ::	
Db	156	ACCAATTATAGACCACAAAAAAGAAGCTGACACATTAATAAGAAAACCACAAGTGAAGCAAAG	215
Oy	349	AspleuphegluaLnLysLIlgLYcIn-----GLYThrValasp---pheileasnaSn	365
		::::	
Db	216	CTGGAGATAGAAACC---CTGGAGAAAGAAATCTCGAACCCATAGATGGCAGCATTCAGCAAC	272
Oy	366	GluIleatgaspProSerLysAlaLeuLleatrgLysValSerthrGIyaLagLuaspLeu	385
		:      :    :	
Db	273	AGAAATACAGAGATGGAAGAGATC-----TCAGTCTACGAAGATGCC	317

Oy	386	Phegluasn 		-lysllleqlnglyl	393
Dd	318	AAGAGAACATCTGCACACACGAAAGAATAATGCAAAAGATCCTACTAA			
Oy	394	Thrvlaaspheilleasnslgllleargsproserylalauilelrqylval			413
Dd	378	AACATCCAGTAATGCGAGCAACAATGGAAAGACCNAAT-----		---CACCGATTATA	428
Oy	414	TyrThrctualaaaspaapleupheglunslsllleqlnglylthrvalaspheile			433
Dd	429	GCAGTAGTGTGAATGAAGATTTCACCTTAA-----		---GGCCAGCAACCTTTCTC	479

```
QY      434 AsnLysGluIleArgAsp 4339
          ||||| |||
Db      480 AACCAAAATTATGAGAA 497
```

RESULT 15		
LOCUS		
A2152840	324 bp	DNA
A2152840		linear
SP_0034_B2_G06.SP6E	Strongylocentrotus purpuratus,	GSS 28-AUG-2000
DEFINITION	purpuratus, purple sea	

urichin, sperm genomic	
genomic clone Plate=34	
AZ152840	
AZ152840.1	GI:8304741

KEYWORDS	SOURCE	ORGANISM
GSS.	Strongylocentrotus purpuratus.	
	Strongylocentrotus purpuratus	
	Euarystea; Metazoa; Echinodermata;	Eleutherozoa; Echinozoa;
	Echinoidea; Euechinozoa; Echinacea; Echinoida;	

## REFERENCE AUTHORS

Strongylocentrotidae; Strongylocentrotus.  
1 (bases 1 to 324)  
Cameron, R.A., Mahlars, G., Rast, J.P., Martinez, P., Bondi, T.R.,  
Swartzell, S., Wallace, J.C., Postka, A.J., Livingston, B.T., Way,  
G.A., Ettenson, C.A., Lehrsch, H., Britten, R.J., Davidson, E.H. and  
Hood, L.

TITLE	A sea urchin genome project: Sequence scan, virtual map, and
JOURNAL	additional resources
Proc. Natl. Acad. Sci. U. S. A.	97 (17), 9514-9518 (2000)
COMMENT	20402566
CONTACT	Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421

FEATURES

SOURCE

Location/Qualifiers
1. .324

BASE COUNT	81	a	95	c	91	g	57	t
ORIGIN	/organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="plate=34 Col=12 Row=N" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: Sperm; Vector: BACs.6; BAC clones in E-Coli DH10B"							

Alignment Scores:	
Pred. No.:	0.0135
Score:	109.00
Percent Similarity:	45.13%
Best Local Similarity:	26.55%
Query Match:	4.62%
DB:	12
	3
	Gaps:
	12
	Indels:
	12
	Mismatches:
	50
	Matches:
	32
	Length:
	328

US-09-807-459-2 (1-458) X AZ152840 (1-324)

QY	294	AsnLysGsluIleargGluProSerLysAlaLeuLysGluLysValSerThrAspThrLys	313
Db	17	AACCAAGCAGCATGAGACTCCCAAGTCACGCCCTGGCTGTCTGGCATGGACAGCCGA	76
QY	314	AspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheAsnLysGluIleArg	333
Db	77	GATTGCATGGGAACCCGGTATTCTTTGGGGGAT-----AACCAAGCCATCAAG	124
QY	334	AspSerLeuLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsn	353
Db	125	ACGTCAAGTCAACCTCGGCCGTGTCTGTGGCAGGGGACACCCAGAATGATGGAGAAC	184
QY	354	LysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleargAspProSerLysAla	373
Db	185	CGGATATCTTCGGGGGAC-----AACCAAGCAGCATCAAAACCCCAAGTCAGCC	232

QY 374 LeuIleArgLysValSerThGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGly 3933  
Db 233 TCGGCCGCTGCTCGGCATGGGACAGCCGAGAACCGATGGGAGAACCGATATTCCTGGGG 2923

Oy	394	ThrValAspPheIleAsnAsnGluIleArgAspProSer	406
Db	293	GAT-----AACCAGCCATCGAGACCCCAAGT	319

RESULT	16
LOCUS	AZ988471
DEFINITION	642 bp DNA linear GSS 27-APR-2001
ACCESSION	AZ988471
	2M0271021F Mouse 10kb plasmid U06C2M library Mus musculus genomic clone U06C2M0271021 F, DNA sequence.
	AZ988471

VERSION A2988471.1 GI:13859698  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 REFERENCE 1 (bases 1 to 642)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weis, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0271 Row: 0 Column: 21  
 Seq primer: CATTGTAAACGACGCCACT  
 Class: plasmid ends  
 High quality sequence stop: 642.  
 Location/Qualifiers  
 1. 642  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUCG2M0271021"  
 /clone\_lib="Mouse 10kb plasmid UUCG2M library"  
 /sex="Female"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv. Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 286 a 137 c 116 g 103 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.0391 length: 642  
 Score: 109.00 Matches: 61  
 Percent Similarity: 41.77% Conservative: 43  
 Best Local Similarity: 24.50% Mismatches: 77  
 Query Match: 4.62% Indels: 68  
 DB: 12 Gaps: 14  
 US-09-807-459-2 (1-458) x A2988471 (1-642)  
 QY 225 GlnHisIleTherSerSerrTyLysAspPtyrMetAspThrGlnIleProAlaLeuProLys 244  
 Db 19 AAGAAATCTTACTTAACGAACCAACGACCTACACATCATCAGAACCCAGACATCC-- 75  
 QY 245 PheAlaLysArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAlaGlyTyr 264

Db 76 -----ACTTGCCAGTCCAGGGCAC 96  
 QY 265 ValAspThrProTyrTyLysTyrMetLysLeuLysAsnPheMetValAsnArg 284  
 Db 97 CCCAAACACCC-----GAAACCTATACCCGGATTAAAGCCTATCTCATG---ATG 147  
 QY 285 ValPheIleProThrLysLysPhePheAsn-----LysGluIleArgGluProSer 301  
 Db 148 ATGTTAAGAGCATCAAGAGGACATTAAATCACTTAAAGAAATACAGAGAACACT 207  
 QY 302 -----LysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGlu 317  
 Db 208 GCTAAATAGTAGAAGACATTAAAGAGAGAACCAAAATCCCTTAAGAAATTTCACAGAA 267  
 QY 318 AsnLysIleGlyGlnGlyThrValAspPheAsnLysGluIleArgAspProSerLys 337  
 Db 268 AACCAACCAACAGC-----GTGATGGAATTGAAATTAACCATCAACAGC----- 312  
 QY 338 AlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGlu----- 352  
 Db 313 --CTAAAGGAGGAAGTAGACACCAATTAAGAAACCCAAAGTGAGCAACGCTGAGATA 369  
 QY 353 -----AsnLysIleGlyGln---GlyThrValAsp---PheIleAsnArgGluIle 367  
 Db 370 GAAACCTAGGACACCATAGCAAAATCTGGAACCTACATGCGACATCAGCAACAGATA 429  
 QY 368 ArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGlu 387  
 Db 430 CAACAGATGCAAGAGACAGATC-----TCAGGTGAGAAAGANTTCATGAGAA 474  
 QY 388 AsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLys 407  
 Db 475 AAC-----ATCGGACCAACCAATCAAAAGAAATATACAAA 507  
 QY 408 AlaLeuIleArgLysValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGln 427  
 Db 508 TCC-----AAAAGATCTCTAATCTCAA----- 528  
 QY 428 GlyThrValAspPheIleAsnLysGluIleArgAspProSerLysAlaLeuIleArgLys 447  
 Db 529 ---AACATCCAGGAATTCAGACACATGAGACCAATGCAACCAATGATATACGAGTA 585  
 QY 448 ValSerThrGluAlaAspAsnLeuLeu 456  
 Db 586 GATGAATGAAATGAAATTTTCGACATCATTA 612  
 RESULT 17  
 BB230024 700 bp mRNA linear EST 23-OCT-2001  
 LOCUS BB230024  
 DEFINITION BB230024 RIKEN full-length enriched, 3 days neonate thymus Mus  
 ACCESSION BB230024  
 VERSION BB230024.2 GI:16354349  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota: Metazoa: Chordata: Cranialia: Vertebrata: Euteleostomi:  
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
 REFERENCE 1 (bases 1 to 700)  
 AUTHORS Arakawa, T., Carlini, P., Fukuda, S., Furuno, M., Hanagaki, T., Harra, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, C., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
 JOURNAL Unpublished (2001)  
 COMMENT On Jul 3, 2000 this sequence version replaced gi:8900669.  
 Contact: Yoshinori Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute





/note="Vector: PMD42nv: Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (9147321149b|AF19072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 326 a 142 c 134 g 119 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.0536 Length: 721  
Score: 108.50 Matches: 59  
Percent Similarity: 41.56% Conservative: 42  
Best Local Similarity: 24.28% Mismatches: 75  
Query Match: 4.60% Indels: 67  
DB: 12 Gaps: 13

US-09-807-459-2 (1-458) x AZ814810 (1-721)

OY 223 ArgLeuGlnHisIleIleTherSerTyLysAspTyrMetAspThrGlnIleProAlaLeu 242  
DB 49 AAATGTAGAACTTACTTACAGAAAAACAGACACATCATCATCAGAACTTACACT 108  
OY 243 ProlysPheAlaLysArgPheSerLeuMetValValGlnArgLeuAlaThrValAla 262  
DB 109 CCCACCTCAGCCAGT-----CCG 126  
OY 263 GlyTyrValAspThrProTrrPtyrLysTyrMetLysLeuLysAsnPhMetVal 282  
DB 127 GGAACCCCAACACACT-----GAAAGCTAGACCTGATTTAAACCATTTCTCAG 180  
OY 283 AsnArgValPheIleProThrLysLysPheAsn-----LysGluIleArgGluPro 300  
DB 181 ATGATGCTAGAGAACATCAGAGGAGCTTTAAGAACTCATTAAGAAATACAGGAGAAC 240  
OY 301 Ser-----LysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPhe 316  
DB 241 ACTGCTACAGAGTACAACTCTTAAGAAAAACAGAAACACAAACCGGTGATG 300  
OY 317 GluAsnLysIleGlyGlnGlyThrValAspPheAsnLysGluIleArgAspProSer 336  
DB 301 GAA-----TTGAAACAATATCATACAGAC----- 324  
OY 337 LysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGlu----- 352  
DB 325 -----CTAAAGAGGAAAGTACAGACATTAAGAAACCCCAAGTAGAGCAACGCTGGAG 378  
OY 353 ---AsnLysIleGlyGln-----GlyThrValAspPhe---IleAsnAsnGluIleArg 368  
DB 379 ATAGAAACCTTGAAAGAAATCTGGAACCATAGATATGAGCATTTAACAACAGAAATACAA 438  
OY 369 AspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsn 388  
DB 439 GAGATGGAAGAAAGATC-----GCAGGTGCGAGAAATTCATAGAGGAGAC 483  
OY 389 -----LysIleGlyGlnGlyThrValAsp 396  
DB 484 ATCAGCACACAAATCAAAAGAAATACAAATGCAAAAGATCTTAACTCAAAACATCCAG 543  
OY 397 PheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGlu 416

DB 544 GAAATCCAGAGACACAAATGAAAGACCAAC-----CTACGGAATATAGAGTAGAT 594  
OY 417 AlaAspAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnLysGlu 436  
DB 595 GAGAAATGAAGATTTTGCATTTAA-----GGGCGCACTTAATATCTTCAACAAAAAT 645  
OY 437 IleArgAsp 439  
DB 646 ATGAAAGAA 654

RESULT 20  
CNS06HIG/c 996 bp DNA linear GSS 17-JUN-2001  
LOCUS T3 end of clone AS0AA012F08 of library AS0AA from strain CLIB 533  
DEFINITION of Saccharomyces bayanus, genomic survey sequence.  
ACCESSION AL399038  
VERSION AL399038.1 GI:12153033  
KEYWORDS GSS.  
SOURCE Saccharomyces bayanus.  
ORGANISM Saccharomyces bayanus  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 996)  
AUTHORS Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P., Aigle,M. and Durrens,P.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 5. Saccharomyces bayanus var. uvarum  
JOURNAL FEBS Lett. 487 (1), 37-41 (2000)  
MEDLINE 20584715  
REFERENCE 2 (bases 1 to 996)  
AUTHORS Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S., de-Montigny,J., Dijon,B., Durrens,P., Leplingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozler-Kalotopoulou,O., Potter,S., Saurin,W., Tekala,F., Toffano-Mloche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
MEDLINE 20584711  
REFERENCE 3 (bases 1 to 996)  
AUTHORS Direct Submission  
TITLE Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.  
LOCATION/Qualifiers  
1..996

FEATURES  
source  
/organism="Saccharomyces bayanus"  
/strain="CLIB 533"  
/variety="uvarum"  
/db\_xref="taxon:4931"  
/clone="AS0AA012F08"  
/clone\_11b="AS0AA"  
/note="end : T3"  
complement(4..>939)  
/note="similar to Saccharomyces cerevisiae ORF YFL008w [ SMC1 : chromosome segregation protein ]"  
evidence=not\_experimental  
misc\_feature  
BASE COUNT 233 a 212 c 155 g 395 t 1 others  
ORIGIN







```

Oy 212 Prolenasp-----TleclYthrgluHsSerVal-----SerargleuGlnHisIle 227
    |||
    |||
    |||
Db 869 AATAAGACACAAACGCTGAATAAATGAGATTCGTACAGACAAATGAGAGTTGAAGGATA 810
Oy 228 ThirSerSerTyLysAspTyMetAspThrGlnIleProAlaLeuProLysPheAlaLys 247
    |||
    |||
    |||
Db 809 AGTTCAGAAATGAGAGAG-----792
Oy 248 ArgPheSerLeuMetValValGlnArgleuLeuAlaThrValAlaGlyTyrrValAspThr 267
    |||
    |||
    |||
Db 791 -----TTGAAAAGCATTAATAGTATTATAAGAAAGCAATGATCATCATATAACA 741
Oy 268 ProTrrTyrrLys-----LysTrrTyrrMetLysLeu-----277
    |||
    |||
    |||
Db 740 AACTATGATTAAGTAGTAAGAGAGAAATTACAGTATTCGATTAACATAGAGAACTTGAT 681
Oy 278 LysAsnPheMetValAsnArgValPheIleProThrLysLysPhePheAsnLysGluIle 297
    |||
    |||
    |||
Db 680 GAAAACGACTTGATTAATTCCTTCGTATCAAGAAATTCGACACATCAACAACAAATG 621
Oy 298 ArgGluProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGlu 317
    |||
    |||
    |||
Db 620 AACGAAAGCGTAATRAATTTTCAAAAAATTAATGACACAACTGTAATTAAGTTAAATGAG 561
Oy 318 AsnLysIleGlyGlnGlyThrVal---AspPhePheAsnLysGluIleArgAspProSer 336
    |||
    |||
    |||
Db 560 TTAGAAATTAAGCAAAACCAATCTGAAGCCTTACTTTCAGAAAGAAATTAAGGAGGAAAT 501
Oy 337 LysAlaLeuLysGluLysValSerAsnAlaLysAspLeuPheGluAsnLysIleGly 356
    |||
    |||
    |||
Db 500 GAAACTCTAAACCAACAAATTACAGAAAGCGAAACGATTTGCTC-----456
Oy 357 GlnGlyThrValAspPheLeuAsnGlnIleArgAspProSerLysAlaLeuIleArg 376
    |||
    |||
    |||
Db 455 -----AGAAATTCGAACCTCCAGACAGATGAGTTACTATCT 423
Oy 377 LysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp 396
    |||
    |||
    |||
Db 422 AAACAAACAAATCTGAAGCTGAT---TTTGAAATTCAAACTACTACTATCATTAATTAAT 366
Oy 397 PheLeuAsnAsnGlnIleArgAspProSerLysAlaLeuIleArgLysValTyrrThrGlu 416
    |||
    |||
    |||
Db 365 AAATGAAATAGGCTCTGATGAAGAATTAATGCTTGAAAGCAAGCAATGAAACT 306
Oy 417 AlaAsp-----AspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp 431
    |||
    |||
    |||
Db 305 GGTGATTAATAGTAGATTAAGCAATTAATTCGAAGATGATGATTCAA---CGATTTAAC 249
Oy 432 PheIleAsnLysGluIleArgAspProSerLysAla-----443
    |||
    |||
    |||
Db 248 ATGTTAAATAGCGAAATCAAGAAATAGACCAAGCATTTCAAGAAACGAGAGAAACTCG 189
Oy 444 LeuIleArgLysValSer---ThrgluAlaAspAsnLeuLeuGluLys 458
    |||
    |||
    |||
Db 188 CTATCTAGATTAATGAGTGTACTGACCAAGAAATATGCTTAAAAAG 141

```

## RESULT 22

```

A2431579 642 bp DNA linear GSS 03-OCT-2000
LOCUS 1M0216006R Mouse 10kb plasmid UUCG1M library Mus musculus genomic
DEFINITION clone UUCG1M0216006 R, DNA sequence.
ACCESSION A2431579
VERSION A2431579.1 GI:10555592
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.

```

REFERENCE 1 (bases 1 to 642)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 JOURNAL plasmid inserts  
 COMMENT Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 1000 Std Error: 0.00  
 Plate: 0216 row: 0 column: 06  
 Seq primer: CACACGGAAGAACAGCATATCAC  
 Class: plasmid ends  
 High quality sequence stop: 642.  
 Location/Qualifiers  
 1..642  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUCG1M0216006"  
 /clone\_lib="Mouse 10kb plasmid library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="vector: PMD42nv. Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pMD42 (g1473211419b1AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 294 a 125 c 116 g 107 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.0585 Length: 642  
 Score: 107.50 Matches: 58  
 Percent Similarity: 44.78% Conservatve: 32  
 Best Local Similarity: 28.86% Mismatches: 66  
 Query Match: 4.56% Indels: 45  
 DB: 12 Gaps: 12

US-09-807-459-2 (1-458) x A2431579 (1-642)

```

Oy 263 glyTyrrValAspThrProTrrTyrrLysLysTyrrMetLysLeuLysAsnPheMetVal 282
    |||
    |||
    |||
Db 96 GGGCACCACCAACACACC-----GAAAAGATVAGACCCGAGATTAAAGCATATCTCAG 149
Oy 283 AsnArgValPheIleProThrLysLysPhePheAsn-----LysGluIleArgGluPro 300
    |||
    |||
    |||
Db 150 ATGATGATAGAGAGACATCAAGAACTTAACTCACTTAAGAAATATACGAGAGAGAC 209
Oy 301 Ser-----LysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPhe 316
    |||
    |||
    |||
Db 210 ACTGCTAAACAGATAGCAATCTTAAGAGGAAAGCAAGCAAAATCCCTTAAAGATTTCAG 269
Oy 317 GluAsnLysIleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSer 336
    |||
    |||
    |||
Db 270 GAAACATGACCAACACAG-----GTGATGGAATTAATTAACCAACCAAGAC-----317

```

QY	337	LysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGlu-----	352
Db	318	-----CTAAAAAGAAAAGTAGATCAATTAAGAAAAAACCAACTGAGCAACTGCAG	371
QY	353	--AsnLysIleGlyGlu-----GlyThrValAspPhe-----IleAsnAsnGluIleArg	368
		:::::           :::	
Db	372	ATAGAAACTTAGAAAGCAATCTGTGCACCATAGATGCGAAATATCAGCAACAGGATACAA	431
QY	369	AspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAsn	388
		::: ::: :::::     :::	
Db	432	CGCATGGAAGAGAGATC-----TCAGGTGCAGGAAGATTCCATACAGAAAT	476
QY	389	-----LysIleGlyGlnGlyThrValAsp	396
Db	477	ATCTCCACACACAGTCAAAAGAAAATACAAAAATGCCAAATTCGAACTCAAAATATTCGAG	536
QY	397	PheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyThrGlu	416
		:::::           :::	
Db	537	GAAATCCAGACACACAAATGAGAAAGCCCAAAACCTAGTGATATAT-----GAA	581
QY	417	AlaAspAspLeuPheGluAsnLysLysIleGlyGlnGlyThrValAspPheIleAsnLysGlu	436
		::: ::: :::::     ::: :::	
Db	582	GTAGATGAGAAATGAAATTTTCACACTT-----AAAGGAAACAACAAAATATCTTCAACAAATT	638
QY	437	Ile	437
Db	639	ATA	641

RESULT 23	LOCUS	DEFINITION
AZ733090	640 bp	DNA linear
RPC1-24-149N17	Mus musculus	GSS 25-JAN-2001
RPC1-24-149N17	genomic clone	RPC1-24-149N17
DNA sequence.		

ACCESSION	AZ/33090	
VERSION	AZ733090.1	GI:12496789
KEYWORDS	GSS.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE  
1 (bases 1 to 640)  
AUTHORS  
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M.,

JOURNAL	Mouse BAC End Sequences from Library RPCI-24
TITLE	Unpublished (1999)
COMMENT	Other_GSSS: RPCI-24-149N17.TV

Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [szhaoc@tigr.org](mailto:szhaoc@tigr.org)  
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong ([pejong@mai.cho.org](mailto:pejong@mai.cho.org)). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end plate: [http://www.tigr.org/tdb/bac-ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html)  
Plate: 149 row: N column: 17  
Seq primer: Sp6  
Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .640

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-149N17"
/clone_id="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pYARAC1; Site_1: BamHI; Site_2: BamHI.
RPCI-24 Mouse BAC Library produced by Pieter de Jong.

```

```

library was cloned in the pTARBAC1 cloning vector at the
Bam1 sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      281 a      126 c      129 g      104 t
ORIGIN
Alignment Scores:
Pred. No.:      0.0869      Length:      640
Score:          106.00      Matches:      63
Percent Similarity: 42.68%      Conservative: 39
Best Local Similarity: 26.36%      Mismatches:  83
Query Match:     4.49%      Indels:      55
DB:              12      Gaps:        13

```

US-09-807-459-2 (1-458) x A2733090 (1-640)

QY	191	ThrTrpGlyLeuPheGlyTrp-----GlyIleAspGluIleGluIle	207
		:::	
Db	28	AcAcCAGAGATCAGAGATGGACAGGCTCCAGTCAGACAAAGCAGGCGAGTATCACT	87
QY	208	ArgSerAsnLeuProLeuAspIleGlyThrGluHisSerValSerArgLeuGluHisIle	227
		:::	
Db	88	AcAGATACACAGATGGCGGAGTCAAGTGTGAAGCAATTATCAACAGAACCA-----	141
QY	228	ThrSerSerTyrLysAspTyrMetAspThrGluIleProAlaLeuProLysPheAlaLys	247
		:::	
Db	142	-----ATTACTTGGCAATCAGAACCAAGATTCTCC-ATTGGCAAGT	185

OY 248 ArgPheSerLeuMetValValGlnArgLeuLeuAlaThrValAlaGlyTyrValAspThr 267

Db 186 -----CCTGGATACTCCACACA 203

QY 268 PROTPYRLLYSYSTPTPYRmettLysLeuLysasnPhmetValasnArgValpheile 287

Db 204 CCA-----GAAAGCAAGATATAGATCGAAAGTCACTTCTCATGATGAGATCAGAGGTA 257

QY 288 ProThrLysLysPheAsn-----LysGluIleArgGluProSer----- 301

Db 258 CTTAAGAGGACATAATACTCCCTTAAGAAATACAGGAGAACACAGGTAACACAGCTA 317

QY 302. LysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGly 321

Db 318 GAAGCACTTAAGAGAGAAACACAAAATCCCTTAAGAATTACAGGAAACACAAATCAA 377

QY 322 GINGLYThrValAspPheAsnLysGluIleArgAspProSerLysAlaLeuLysGlu 341

Db 378 CAG-----GTGAAGGAAGGACAAACATCCAGATCTAAAAATGGAATAGAACA 431

QY 342 ---LysValSerAsnSpAlaLysAspLeuPheGluAsnLysIleGlyGln-----Gly 358

Db 432 ATAAGAATCACAGAGGACACCCTGTAGTTAGAAACCTAGGAAGAGATCAGGA 491

QY 359 ThrValasp---PheIleasnAsnGluIleargaspProSerLysAlaLeuIleargLys 377

Db 492 GTCATAGATGCAAGCATCGCCACACAGAATACAAGAG-----ATAGAAAG 536

QY 378 ValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhe 397

Db 537 AGAATCTCAGGTGCAGAGATACCTTAGAAAC----- 569

QY 398 ILEASNGLUILLEARGASPROSERLYSALALEUILLEARGLYSVALTYRTHRGLU 416

Db 570 ATGACACA CTGTCAAGAAATGCAAAAGAC-----AAAAAGCTCCTAACCCAA 620

RESULT 24

LOCUS	AZ462925	478 bp	DNA	linear	GSS 04-OCT-2000
-------	----------	--------	-----	--------	-----------------

clone UUGC1M0271A02 R, DNA sequence.

VERSION AZ462925.1 GI:10621050

SOURCE house mouse.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 478)			
1	Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts	Unpublished (2000)	Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: adunn@genetics.utah.edu Insert length: 10000 Std Error: 0.00 Plate: 0271 row: A column: 02 Seq primer: CACACAGCAACACCTATGACC Class: plasmid ends High quality sequence stop: 478. Location/Qualifiers
1	.478			
1	/organism="Mus musculus"			
1	/strain="C57BL/6J"			
1	/db_xref="taxon:10090"			
1	/clone="U02C1M0271A02"			
1	/clone_id="Mouse 10kb plasmid U02C1M library"			
1	/sex="Male"			
1	/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"			
1	/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g14732114[gblAF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."			
BASE COUNT	100 a	95 c	92 g	191 t
ORIGIN				
Alignment Scores:				
Pred. NO.:	0.0721	Length:	478	
Score:	105.00	Matches:	46	
Percent Similarity:	44.38%	Conservative:	33	
Best local Similarity:	25.84%	Mismatches:	55	
Query Match:	4.45%	Indels:	44	
DB:	12	Gaps:	8	
US-09-807-459-2 (1-458) x AZ462925 (1-478)				
OY	197 TTPGIIILYLSARGALAEULYSGINLEILAEYSGERRANLEUPLAEUSPLIEG	216		
	:::    :::            :::         :::			
DB	441 TTTGGTGTCCAGAGCGTTTGACACAGGTTTTCANA-----CTTCTATAGCT	394		
OY	217 ThrGluHISerValSerArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetasp	236		
	:::        :::     :::			
DB	393 ACAAGCTCCACGACAGTCAGTATTATATCATCATCAACACCCAGAGATGACCAACAAATGCTGAAG	334		

Oy	237	ThrgInIleProAlaeLeuPolysPhehalalys-----ArgPheSerIeu	251
Db	333	GCAACGTAAGAACATCTTACCAGAAACAAGACTACATGGCATCATCACCACTA	274
Oy	252	MetValValGlnArGLeuLeuAlathrValAlacIlyThrValAspThrProTrpIlys	271
Db	273	CTTTACACAGCCAGTCCTC-----GATACCCCAACATACCGG	235
Oy	272	LysTrpIyrMetLys--LeuLysasnPhemetalAsnArGValPheIleProthrIysl	291
Db	234	AAAAGCAAGATTCCAGATTAAATCATATCAATGATGCTGTGA-----	191
Oy	291	ysPhePheAsnLysGluIleArGluProSerLysAlaLeuLysGluLysValserThra	311
Db	190	-----GAGAAATTAAAGCAACATTAATAATT	163
Oy	311	sPTThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheAsnLysG	331
Db	162	GCCTTAAAGAATACAGAGAACACAGTAAACAGGTGAAAGTC-----GCTAAAG	112
Oy	331	IuileArGspProSerLysAlaLeuLysGluLysValserAsnPalalysAspleuP	351
Db	111	AG-----GAGGACAAATAATCCCTTAAAGATTACAGAAACACACAAACGACGTGA	58
Oy	351	heglu---AsnLysIleelYglnGlyThrValAspPheIleAsnslcu	366
Db	57	AGGATTTCACAAAA-----ACCATCCAGTACTGTGAAGAAG	20
RESULT 25			
BMT273998			
LOCUS	BM273998	532 bp	mRNA linear EST 20-DEC-2001
DEFINITION	PfESToa62d10.y1 Plasmodium falciiparum 3D7 gametocyte cDNA library		
ACCESSION	Plasmodium falciiparum 3D7 cDNA 5' mRNA sequence.		
VERSION	BM273998		
KEYWORDS	BM273998.1 GI:17967300		
SOURCE	EST.		
ORGANISM	Plasmodium falciiparum 3D7.		
TITLE	Plasmodium falciiparum 3D7		
REFERENCE	Eukaryota: Alveolata: Apicomplexa: Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 532) Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Matta,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritzer,E., Bennett,J., Jentes,E., Konko,I., Tsagarelshvili,I.R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Richey,J., Madkins,J., Kennedy,S., Levinso,D., Waterston,R., Wilson,R., and Sibley,D. Washu Plasmodium EST Project Unpublished (2001)		
JOURNAL COMMENT	Contact: L. David Sibley Washu Plasmodium EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wuston.wustl.edu Library was constructed by R. Haywood. DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: L. David Sibley (sibley@orcin.wustl.edu), Washington University Seq primer: -40up from GIBCO High quality sequence stop: 422. Location/Qualifiers 1..532 /organism="Plasmodium falciiparum 3D7" /db_xref="taxon:36329" /clone_lib="Plasmodium falciiparum 3D7 gametocyte CDNA library" /dev_stage="gametocyte (stage III-V)" /lab_host="DH10B (Genescript Invitrogen, Inc.)" /note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2: XhoI; The library was constructed by R Haywood. cDNAs were synthesized from gametocyte poly(A)+ RNA by oligo d(T) priming, size-selected and directionally cloned into the		
FEATURES			
SOURCE			



QY 302 Lysala-----LeuLysGluLysValSerThrAspThrLysAspLeu 315  
 DB 318 AACACTGCTAAACAGGTAGACAGCATTTAAAGGAGCAACAAAATCCCTTAAAGATTG 259  
 QY 316 PheGluAsnLysIleGlyGlnGlyThrValAspPheAsnLysGluLysGluLeuArgAspPro 335  
 DB 258 GAGGAAACACACCAACCAACAG-----GTGATGGAATTCATTAATTCATCCAAAGAC--- 208  
 QY 336 SerLysAlaLeuLysGluLysValSerAsnAspAlaLys----- 348  
 DB 207 -----CTAAAGAGGAGTAGACACAAATTAAGAAAGAACCCAAAGTGAGGCAACGCTG 157  
 QY 349 -----AspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp---PheIleAsnAsn 365  
 DB 156 GAGATAGGAACCTAGACAAAGAAA-----TCTGGAACCATAGATCGAGCATCAGCAAC 103  
 QY 366 GluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeu 385  
 DB 102 AGAATACAGAGATGAGAGAGATC-----TCAGTGTGAGAAAGATTCC 58  
 QY 386 PheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspPro 405  
 DB 57 ATAGAGAAC-----ATCGGCAACAACATCAAAAGAAAG 25  
 QY 406 SerLys 407  
 DB 24 GAAAG 19

RESULT 27  
 LOCUS BM168451 644 bp mRNA linear EST 04-DEC-2001  
 DEFINITION ESF570974 PYBS Plasmodium yoelii yoelii cDNA clone pYCPA68 5' end,  
 mRNA sequence.

ACCESSION BM168451 GI:17301683  
 VERSION BM168451.1  
 KEYWORDS EST.  
 SOURCE Plasmodium yoelii yoelii.  
 ORGANISM Plasmodium yoelii yoelii.  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 644)  
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,  
 Fraser,C.M. and Carucci,D.J.  
 TITLE Plasmodium yoelii EST project at TIGR  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Jane Carlton  
 Parasite Genomics Group  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-530-9319  
 Fax: 301-838-0208  
 Email: carlton@tigr.org  
 For clone info, please contact the Malaria Research and Reference  
 Reagent Resource Center, ATCC  
 http://www.malaria.mr4.org/mr4pages/index.html  
 Seq primer: ADF.

FEATURES  
 source Location/Qualifiers  
 1..644  
 /organism="Plasmodium yoelii yoelii"  
 /strain="17XL"  
 /db\_xref="taxon:73239"  
 /clone="PYCPA68"  
 /clone\_1lb="PYBS"  
 /dev\_stage="Asexual blood stages"  
 /lab\_host="E. coli XL-1 Blue"  
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was  
 collected from BALB/cByJ mice infected with Py17XL  
 parasites, and leukocytes removed by passage over  
 microcrystalline cellulose columns. Total RNA was  
 isolated using the guanidium isothiocyanate method, and  
 mRNA isolated using oligo(dT)-cellulose chromatography.  
 First strand cDNA synthesis was completed using a 50-base  
 primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven  
 termini were treated with Pfu DNA polymerase and EcoRI  
 adaptors ligated to the blunt ends. The sample was cleaved  
 with XhoI and separated on a Sephacryl S-500 column.  
 Size-fractionated cDNA was precipitated and ligated to  
 HybridZAP arms directionally using EcoRI-XhoI cleaved arms.  
 After packaging, the phagemid vector (PAD-GAL4) was  
 excised from the HybridZAP vector and plasmid DNA  
 isolated."

BASE COUNT 310 a 69 c 109 g 156 t  
 ORIGIN

#### Alignment Scores:

Pred. No.: 0.15 Length: 644  
 Score: 104.00 Matches: 48  
 Percent Similarity: 46.328 Conservative: 40  
 Best Local Similarity: 25.268 Mismatches: 66  
 Query Match: 4.41% Indels: 36  
 DB: 10 Gaps: 11

US-09-807-459-2 (1-458) x BM168451 (1-644)

QY 276 LysLeuLysAsnDheMetValAsnArg-----ValPheIlePro 288  
 DB 37 AAATATATAAATATGAGTAGATTAACCGAACAAATTAATAATGACATGCAATTA 96  
 QY 289 ThrLysLysPhePheAsnLysGluIleArgGluProSerLysAlaLeuLysGluLysVal 308  
 DB 97 ACCGAAATAATATAAATAT--GAAGTATCGAATTAACCGAAATAATATAAATATGAAGTA 153  
 QY 309 SerThrAspThrLysAspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPhePhe 328  
 DB 154 TCGAATTAACCGAAATAATTA--AAATGAGATTCGATTAACCGAAATAATTAATA 210  
 QY 329 AsnLysGluIleArgAspProSerLysAlaLeuLysGluLysValSerLysAlaLys 348  
 DB 211 AAT---GAAGTACCTAATTAACCGAACCAATTAATAACACAGCTA---GATGATATTA 264  
 QY 349 AspLeuPheGlu-----AsnLysIleGlyGlnGlyThrValAspPheIleAsnAsn 365  
 DB 265 GATTAATATTAAACATAGCTGTAACACTCTTGATGAGATCAATCCCAATCGAATAAT 324  
 QY 366 GluIleArgAspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGlu----- 383  
 DB 325 ACTATTAAACATACCTGTTGGTACATAGGGGATATGCTCTGCTATGAAAAACA 384  
 QY 384 ---AspLeuPheGluAsnLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIle 402  
 DB 385 TCTGATGTAGTTAAAGAAAAAGTA-----ATTGAACATTAATTTAATCT--- 429  
 QY 403 ArgAspProSerLysAlaLeuIleArgLysValIleThrGluAlaAspLeuPheGlu 422  
 DB 430 ---GATATTGTCCTCCGCCCTTAAGAGAGATTGT----- 462  
 QY 423 AsnLysIleGlyGlnGlyThrValAspPheIleAsnLysGluIleArgAsp----- 439  
 DB 463 AATATCTGCTGTAATAATATTACTTCATTAATGAAAAAGATTAATAAGATGGAAGTCC 522  
 QY 440 -----ProSerLysAlaLeuIleArgLys 447  
 DB 523 GATTAAGTGCATAAGCATTTTAAAAA 552  
 RESULT 28  
 LOCUS BH126086/c 715 bp DNA linear GSS 19-JUL-2001  
 DEFINITION RPCI-24-289K13.TJ RPCI-24 Mus musculus genomic clone RPCI-24-289K13  
 , DNA sequence.  
 ACCESSION BH126086  
 VERSION BH126086.1 GI:14969598  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;







Oy	351	phegluanlysileglylgl-----gltthrvalasppne-ileasnsgluillear	368
Dd	333	atagaaaac---cttagaaaagaattctgtgaaccattgatgtccatccacgaagaataca	277
Oy	368	gasproserlysalaleuillearglyvalserrhrcylajagiuaspleuphegluas	388
Dd	276	agagatggaaagacaatc-----tcagstgcagaagaagtccatagagaa	232
Oy	368	nlysiileglyinglythrvalasppheileasnsgluileargasproserlysal	408
Dd	231	c-----atcggcacacaaatcaaaagaaaatgaaaaatg	199
Oy	408	aleuillearglyvaltyrthr-----glualaaspsapleuphegluasly	424
Dd	198	c-----aaaamaagcttcaactcaaaacatccaggaaaatccaggcacacatgagaagacc	145
Oy	424	s-----ilegly---glncltythrvalasppheileasnsgluilleargas	439
Dd	144	aaacctatgatataatgagaatttgatgacactgmaaatTTTT-----CAACTTAAGG	94
Oy	439	proserlysalaleuillearglyvalserrhrgluilaaspsanleu	455
Dd	93	accacgaatatcttcaacacagattatagaaagaaaactttcccAAACCTA	45

RESULT 31	
LOCUS	A2946473/c
DEFINITION	A2946473 439 bp DNA linear GSS 27-APR-2001
ACCESSION	2M02080113
VERSION	clone U00C2M0208013 F, DNA sequence.
KEYWORDS	A2946473
SOURCE	A2946473.1 GI:13815147
ORGANISM	GSS.
REFERENCE	house mouse.
AUTHORS	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
	1 (bases 1 to 439)
	Dunn,D., Aoyagi,A., Barber,M., Baecorn,T., Duval,B., Hamll,C.,
	Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
	,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
	and Wright,D., Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb
	plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss
	University of Utah Genome Center
	University of Utah
	Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
	84112, USA
	Tel: 801 585 5606
	Fax: 801 585 7177
	Email: ddunn@genetics.utah.edu
	Insert Length: 10000 Std Error: 0.00
	Plate: 0208 row: J column: 13
	Seq primer: CGTGTGAACGACGCCACAGT
	Class: plasmid
	High quality sequence stop: 439.
FEATURES	
source	1..439 Location/Qualifiers

**FEATURES**  
**source**

Location/Qualifiers  
1. 439  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="U06C2M0208J13"  
/clone\_lib="Mouse 10kb plasmid U06C2M library"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42uv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/9nares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range, using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (911473211419bIAF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

Alignment Scores:			
pred. No.:	0.123	length:	439
Score:	102.50	Matches:	48
Percent Similarity:	46.88%	Conservative:	27
Best Local Similarity:	30.00%	Mismatches:	40
Query Match:	4.35%	Indels:	45
Db:	12	Gaps:	12

Oy	271	LysLysLysLysLysLysLysLysLysPhe-----MetValAsnArgValPheIle	287
Db	434	GAAGAAGTGAAGAACTGAGATTAAAAACATATCATCATGATGATGTGTAGACACATC-----	381
Oy	288	ProThrLysLysPhePheAsn-----LysGluIleArgGluProSerLysAlaLeu	304
Db	380	-----AAGAAGGAGCTTTAAATACTTACTTTAAAGAAATACAGAACACCTGTAAAGAGTTA	327
Oy	305	-----LysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGly	321
Db	326	CAAGCTCTTAAAGAAAAACAGAAACACATTCACAAACAGGTGATGAA-----	279
Oy	322	GlnGlyThrValAspPhePheAsnLysGluIleArgAspProSerLysAlaLeuLys---	340
Db	278	-----TTTAAACAAACCTACACAGCTTAAAAAAGGAGAGTAGACACA	237
Oy	341	---GluLysValSerAsnAspAlaLysAlaLysAspLeuPheGluAsnLysIleGlyGln-----	357
Db	236	ATTAAGAAACCCCAAAAGAGAGGCAAAAGCTGGAGATAGAAAC--TTGAGAAAGAAATCT	180
Oy	358	GlyThrValAsp---PheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArg	376
Db	179	GGAACCATAGATCGCAGATCATGCACAGACATACAGAGATGGAAGACAGATC-----	126
Oy	377	LysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGlnGlyThrValAsp	396
Db	125	-----TCAGGTGCAGAAAGATCCATAGAGAC-----	99
Oy	397	PheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysValTyrThrGlu	416
Db	98	---ATCGGCAAAACATCAAAAGA-----AAGCGAAATGCAAAAGATCTTAATCTAG	48

RESULT	32
AZ344512/c	
LOCUS	AZ344512 503 bp DNA linear GSS 29-SEP-2000
DEFINITION	JM0078M05R Mouse 10kb plasmid U06C1M library Mus musculus genomic clone U06C1M0078M05 R, DNA sequence.
ACCESSION	AZ344512
VERSION	AZ344512.1 GI:10423749
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 503)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Haml,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingley,A., von Niederhausern,A.



TITLE and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0078 row: M column: 05  
 Seq primer: CACACGAGAAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 503.

FEATURES  
 source Location/Qualifiers

1..503  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Mouse 10kb plasmid UUC1M library"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PMD42 (g14732114|gblaf129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 80 a 101 c 104 g 218 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.152 Length: 503  
 Score: 102.50 Matches: 46  
 Percent Similarity: 46.678 Conservative: 38  
 Best Local Similarity: 25.568 Mismatches: 55  
 Query Match: 4.358 Indels: 41  
 DB: 12 Gaps: 10

US-09-807-459-2 (1-458) x A2344512 (1-503)

OY 223 ArgLeuGlnHisIleThrsSerSerTyrrMetAspThrGlnIleProAlaIeu 242  
 DB 472 AATGTGAGAACTCTATCAACAAACAGACACCTACATCAACACACACATT 413  
 OY 243 ProLysPheAlaLysArgPheSerIleuMetValAlaGlnArgLeuAlaThrValAla 262  
 DB 412 CCC-----ACTTCGCCCACTCA 395  
 OY 263 GlyTyrValAspThrProTrrPtyrLysLysTrrPtyrMetLysLeuLysAsnPhetVal 282  
 DB 394 GGGACACCCCAACACACACC-----AAAAAGCTAGACCTGATTTAAACATATCTCAG 341  
 OY 283 AsnArgValPheIleProThrLysLysPheAsn-----LysCluIleArgGlu 299  
 DB 340 ATGATGTGA---GAGACACCAAGAAAGACTTAAATAACTCACTTAAAGAAATACAGAG 284

OY 300 ProSer-----LysAlaLeuLysGluLysValSerThrAspThrLysAspIeu 315  
 DB 283 AACACTGCTAAACAGGTAGACACATTTAAAGAAAGCAACAAATCCCTTAAAGATTG 224  
 OY 316 PheGluAsnLysIleGlyGlnGlyThrValAspPheAsnLysGluIleArgAsp--- 334  
 DB 223 CAGGAAACACAAACCAACACAG-----GTGATGGAATGTGAATTAACATCCAGACCTTA 170  
 OY 335 -----ProSerLysAlaLeuLysGluLysValSerAsnAspAlaLysAsp 349  
 DB 169 AAAGGGAAGTAGGCACAAATTAAGAAATCCAAAGAGGCAACGCTGATATAGAAAC 110  
 OY 350 LeuPheGluAsnLysIleGlyGlnGlyThrValAsp---PheIleAsnGluIleArg 368  
 DB 109 CTGGAAGAAAGAA-----TCTGGAACCATGATGCGACGATTCGACAAATGATCA 59  
 OY 369 AspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspPheGluAsn 388  
 DB 58 GAGATGGAAGAGGAAATC-----TCAGGTGCAAGAAATTTCAATAGAGAAC 14

RESULT 33 547 bp DNA linear GSS 24-JAN-2001  
 A2696548  
 A2696548/C  
 LOCUS  
 DEFINITION RPCI-23-214L9 TJB RPCI-23 Mus musculus genomic clone RPCI-23-214L9,  
 DNA sequence.  
 ACCESSION A2696548  
 VERSION A2696548.1 GI:12411771  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 547)  
 REFERENCE  
 AUTHORS Zhao, S., Nierman, M., Feldblyum, T., Malek, J., Shatsman, S., Akniet  
 and Fraser, C. M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 Other GSSs: RPCI-23-214L9 TJB  
 COMMENT  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@ligr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@email.cho.org). Clones may be purchased from BACPAC  
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end  
 page: http://www.tigr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 214 row: L column: 9  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers

1..547  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="RPCI-23-214L9"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBAC3.6, site\_1:  
 EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methyase. Size  
 selected DNA was cloned into the pBAC3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 93 a 102 c 109 g 243 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	0.174	Length:	547
Score:	102.50	Matches:	44
Percent Similarity:	46.55%	Conservative:	37
Best Local Similarity:	25.29%	Mismatches:	56
Query Match:	12	Indels:	37
DB:	12	Gaps:	11

US-09-807-459-2 (1-458) x AZ696548 (1-547)

```
OY 223 ArgLeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaLeu 242
      ::::::::::::::::::::
DB 483 CAAATGAGAAATCTTACTAAGAAACCAAGACCAACCTCACCATCATCAGAACCCATCCTACT 424
OY 243 ProLysPheAlaLysArgPheSerLeuMetValGlnArgLeuAlaThrValAla 262
      |||
DB 423 CCCACCTCAGCCACT-----
OY 263 GlyTyrValAspThrProTPrTyrLysLysTPrTyrMetLysLeuLysAsnPheMetVal 282
      |||||
DB 405 GGATACCCCAACACACACC-----GAAAGCAAAACTCGGATTTAAATCATATCTCATG 352
OY 283 AsnArgValPheIleProThrLysPhePheAsnLysGlnIleArgGluProSerLys 302
      |||
DB 351 ATGCTGGA--GTGGACATCAAGAGGGCTTTAATACTCACTTAAGAAATACAGAG 295
OY 303 AlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLysIleGlyGln 322
DB 294 AACACTGCTAAAGATGATGA--GACATTAAGAAATACAGGAAAC----- 250
OY 323 GlyThrValAspPhe-----PheAsnLysGlnIleArgAspProSerLysAlaLeu 339
      |||
DB 249 ---ACAATAAATCGTAATGAATTGAAAGAAACCAATCAAGACCTTAAGAGGAAGTA 193
OY 340 Lys-----GluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGln 357
      ::::::::::::::::::::
DB 192 GAACCAATAAAGAAACCAAGATGATCAACGTTGGATGATGAAGACC--ATAGGAAG 136
OY 358 ---GlyThrValAspPhe--IleAsnAsnGlnIleArgAspProSerLysAlaLeu 374
      |||||
DB 135 AAATCTGGAATATATGATGAGGTCAGCAACAAAGAAATACAGAGATGACAGAGATC 76
OY 375 IleArgLysValSerThrGlyAlaGluAspLeuPheGluAsn 388
      ::::::::::::::::::::
DB 75 -----TCAGGTGACAGAGATTCATAGAGAAC 49

RESULT 34
LOCUS AZ679969 869 bp DNA linear GSS 14-DEC-2000
DEFINITION ENTFF2067R Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION AZ679969
VERSION AZ679969.1 GI:11817115
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 869)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@nih.gov
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
```

Class: shotgun  
High quality sequence start: 15  
High quality sequence stop: 842.  
Location/Qualifiers

1. 869  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"

/db\_xref="taxon:5759"  
/clone\_id="Entamoeba histolytica Sheared DNA"  
/note="Vector: PHOS1. Site 1: Bst I. Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + 1 method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaubin and B.  
Barrell, Oxford University Press, 1999)."

BASE COUNT 366 a 92 c 128 g 283 t

## Alignment Scores:

Pred. No.:	0.408	Length:	869
Score:	102.00	Matches:	47
Percent Similarity:	38.10%	Conservative:	41
Best Local Similarity:	20.35%	Mismatches:	101
Query Match:	4.32%	Indels:	42
DB:	12	Gaps:	10

US-09-807-459-2 (1-458) x AZ679969 (1-869)

```
OY 180 PheGlyAlaSerPhePheAsnLysLeuSerPheThrGlyLeuPheGlyTPrGlyLe 199
      ::::::::::::::::::::
DB 17 TATGCCACTGATTCCTTAATCACTAAGATTAATCACTAATTAATTTGGTTCATTAGTT 76
OY 200 LysArgAlaLeuLysGlnIle----- 206
      ::::::::::::::::::::
DB 77 CAAGCATTAAGAAACCAATAGAGATGAAGTATTATTCATTGAGCCAGTTTATA 136
OY 207 -----IleArgSerAsnLeuProLeuAspIleGlyThrGlnHisSerValSerArg--- 223
      ::::::::::::::::::::
DB 137 GAAGGATTAAGAAATTTGATATGATTCATTCAAAGATGAACCACTTCCACAGATT 196
OY 224 ---LeuGlnHisIleThrSerSerTyrLysAspTyrMetAspThrGlnIleProAlaLeu 242
      |||
DB 197 GATTAAATAGCAATTTCCCAATCATATATATGATTCATTGACGGAACAAAAA 256
OY 243 ProLysPheAlaLysArgPheSerLeuMetValGlnArgLeuAlaThrValAla 262
      |||
DB 257 GATAATTTTACAAAAGATATGACAAATACTAAACAAAGATATGCAACATGTATG 316
OY 263 GlyTyr---ValAspThrPro-----TPrTyrLysLysTPrTyrMetLysLeuLysAsn 279
      ::::::::::::::::::::
DB 317 AAAGTTGGAATGACAAACCAATATATATTCAGATTGGAAGAAAGATGAAATAGCT 376
OY 280 PheMetValAsnArgValPheIleProThrLysPhePheAsnLysGluIleArgGlu 299
      ::::::::::::::::::::
DB 377 ATTTCCATTCATGATTAATTAATTTGATTCATCAAAATTA-----GAGGATTTAAATA 430
OY 300 ProSerLysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPheGluAsnLys 319
      ::::::::::::::::::::
DB 431 GATTACAGAGATTAATTCGAAATTCGAAACCAACCACTTCAATATATACAA----- 484
OY 320 IleGlyGlnGlyThrValAspPhePheAsnLysGluIleArgAspProSerLysAlaLeu 339
      |||
DB 485 -----CAATCATCTAATAGATTTCTTAAT--GAAAGACGGATGATGTC-----ATT 529
OY 340 LysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGlyGlnGlyThr 359
      |||||
```

```

Db      530   AAGAAGACATTATTAATTTGAA-----ACA    559
Oy      360   ValaspheIIeasnaSclulIeargsPProSerLySaLaLeuIlleargLysValSer    379
                |||||  |||  |||  ::|  |||
Db      560   TTTGGATTATTCCCAAAAGTGTTCATCAT-----CAAGTTGCATGAATAATCTCCT    610
Oy      380   ThrGlyAlaGuAspleuPhenclulsnySite    390
                |||  :::::  |||||:::
Db      611   AAATCAGTTGAACAATTATAGTAAACAATAAG    643

RESULT 35
BH070091/c
LOCUS          BH070091              683 bp     DNA             linear    GSS 18-JUL-2001
DEFINITION     RPCI-24-275B18.TV RPCI-24 Mus musculus genomic clone RPCI-24-275B18
               , DNA sequence.
ACCESSION      BH070091
VERSION        BH070091.1   GI:14889688
KEYWORDS       GSS.
SOURCE         house mouse.
ORGANISM       Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 683)
Zhao,S., Nietman,M., Malek,J., Shateman,S., Akintet,B., Levins,M.,
Tsengaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other.GSSes: RPCI-24-275B18.TJ
Contact: Shayng Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pejong@mail.chu.org). Clones may be purchased from BACPAC
Resources (http://www.chori.org/bacpac/orderingframe.html). BAC end
page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 275 row: B column: 18
Seq primer: T7
Class: BAC ends.

FEATURES
            source                  Location/Qualifiers
                                1..683
                                /organism="Mus musculus"
                                /strain="C57BL/6J"
                                /db_xref="taxon:10090"
                                /clone="RPCI-24-275B18"
                                /clone_1lb="RPCI-24"
                                /sex="Male"
                                /cell_type="Spleen/Brain"
                                /note="Vector: pTRABAC1. Site_1: BamHI. Site_2: BamHI.
                                RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                                library was cloned in the pTRABAC1 cloning vector at the
                                BamHI sites using MboI partially digested male C57BL/6J
                                DNA."
BASE COUNT      124 a      130 c      140 g      289 t
ORIGIN
Alignment Scores:
Pred. No.:           0.32                      Length:          683
Score:              101.50                     Matches:         61
Percent Similarity: 42.74%                    Conservative:    42
Best Local Similarity: 25.31%                   Mismatches:     73
Query Match:        4.30%                       Gaps:           66
DB:                 12                          Indels:         13

US-09-807-459-2 (1-458) x BH070091 (1-683)
Oy      190   phenrrrtnGlyleuPheneclyTrppGlyileYsarGaLaLeuLysClnlletleargSer    209
                :::|||| |||||| |||
                :::::

```

[illegible]





QY 319 LysIleGlyGlnGlyThrValAspPheAsnLysGluIleArgAspProSerLysAla 338  
Db 314 -----ATMAACCAACAGGTAGAAGAC----- 294  
QY 339 LeuLysGluLysValSerAsnAspAlaLysAspPheGluIleArgAspProSerLysAla 358  
Db 293 ATTAAGAGGAAACCAAAATCCCTTAAAGATTGCAGAAACCAACCAACAGG--- 237  
QY 359 ThrValAspPheIleAsnAsnGluIleArgAspProSerLysAlaLeu-----IleArg 376  
Db 236 ---GGGATGGATTGATTAATAAACCATCCATAAAGGAGTAGACACATTAAG 180  
QY 377 LysValSerThrGlyAlaGluAspLeuPheGluAsnLysIleGlyGln-----GlyThr 394  
Db 179 AAACCCCAAGTAGAGCAACAGTGCAGATAGAAACA---CTAGGAAGAAATCTGGAAC 123  
QY 395 ValAsp---PheIleAsnAsnGluIleArgAspProSerLysAlaLeuIleArgLysVal 413  
Db 122 ATGATGGCAGCATCGACACAGATACAGAAATGGAAGAGAGCTC----- 75  
QY 414 TyrThrGluAlaAspAspLeuPheGluAsn 423  
Db 74 ---TCAGGTACAGAGATTTCATAGAGAAC 48

RESULT 39  
A2625698 622 bp DNA linear GSS 13-DEC-2000  
LOCUS A2625698  
DEFINITION IM0465N20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0465N20 F, DNA sequence.  
ACCESSION A2625698  
VERSION A2625698.1 GI:11747888  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 622)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rogase, R., Stokes, R., Tinney, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112 USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0465 row: N column: 20  
Seq primer: CCTGTAAAGACGAGCCACGT  
Class: plasmid ends  
High quality sequence stop: 622.  
Location/Qualifiers  
1..622  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0465N20"  
/clone\_id="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PMD42 (g11473211419b) (AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 285 a 116 c 111 g 110 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.317 Length: 622  
Score: 101.00 Matches: 56  
Percent Similarity: 42.93% Conservative: 29  
Best Local Similarity: 28.28% Mismatches: 62  
Query Match: 4.28% Indels: 51  
DB: Gaps: 11

US-09-807-459-2 (1-458) x A2625698 (1-622)

QY 263 GlyTyrValAspThrProTyrLysLysTyrTyrMetLysLeuLysAsnPheMetVal 282  
Db 30 CGATACCCCAACACACACCC-----AAAAGCAAGACCTGAAATTAATCTATCTCATG 83  
QY 283 AsnArgValPheIleProThrLys-LysPhePheAsn-----LysGluIleArgGluPr 300  
Db 84 ATGCTGTAGAGTACTTAAAGAGGCAATTAATTAACCTCTAAAGAAATACAGAGAA 143  
QY 300 cser-----LysAlaLeuLysGluLysValSerThrAspThrLysAspLeuPhe 316  
Db 144 CACTCTTAACACGTAGAGAGTCTTAAAGAGCAACAAAATCTTTAAAGATTACA 203  
QY 316 eGluAsnLysIleGlyGlnGlyThrValAspPheAsnLysGluIleArgAspProse 336  
Db 204 GGAAGAACACCAACCAACAG-----GTAAATGAATTTGAACAAACCATC----- 246  
QY 336 rLysAlaLeuLysGluLysValSerAsnAspAlaLysAspLeuPheGluAsnLysIleGl 356  
Db 247 -CAAAATCTTAAAGAGGAGTAGAACAATAAGAAACCCAAAGTGAACACTCTGA 305  
QY 356 yGlnGlyThrValAsp-----PheIleAsnAsnGluIleArg 368  
Db 306 GATAGGAACCCTAGAAGAAAGAAATCAGAAACATAGTCGAGCATCAACACAGAAATCA 365  
QY 368 gaspProSerLysAlaLeuIleArgLysValSerThrGlyAlaGluAspLeuPheGluAs 388  
Db 366 AGAGATGGAGAGAGCAATC-----TCAGCTCGAGAAAGATTCTATTAGAGAA 410  
QY 388 nLysIleGlyGlnGlyThrValAspPheIleAsnAsnGluIleArgAspProSerLysAl 408  
Db 411 C-----ATGGCACACACA-----ATCAAGATGATGCAAAATG 443  
QY 408 aleuIleArgLysValTyrThrGluAlaAspAspLeuPheGluAsnLysIleGlyGlnGl 428  
Db 444 C-----AAAAAGATTCTAATCA----- 462  
QY 428 yThrValAspPheIleAsnLysGluIleArgAspProSerLysAlaLeuIle 445  
Db 463 -AACATCCAGGAATCCAGACACAAATGAGAAACCAACTAGAGGATATA 513

RESULT 40  
BH313240 700 bp DNA linear GSS 03-DEC-2001  
LOCUS BH313240  
DEFINITION CH230-101L12.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-101L12, DNA sequence.  
ACCESSION BH313240  
VERSION BH313240.1 GI:17242197



**THIS PAGE BLANK (USPTO)**